Strathprints Institutional Repository

Di Maio, Paola (2009) *Common vocabularies for collective intelligence - work in progress.* In: IEEE International Conference on Digital Ecosystems and Technologies (IEEE-DEST 2009), 2009-05-31 - 2009-06-03, Istanbul, Turkey.

http://strathprints.strath.ac.uk/

Di Maio, Paola (2009) Common vocabularies for collective intelligence - work in progress.
In: IEEE International Conference on Digital Ecosystems and Technologies (IEEE-DEST 2009),
31st May - 3rd June 2009, Istanbul, Turkey.

http://strathprints.strath.ac.uk/13047/

Common Vocabularies For Collective Intelligence - Work In Progress

Paola Di Maio, *IEEE Member*

Abstract — **Web based applications and tools offer a great potential to increase the effectiveness of information flow and improve communication among disparate agents in support of distributed operations . One of the factors that hinders the integration and interoperability of information models, is a lack of common, shared vocabularies. This paper discusses research aimed at placing shared vocabularies at the heart of collective intelligence domain, and reports about work in progress of an effort being incubated toward the development of a metadata set aimed at mapping and unifying different vocabularies used in emergency management, and furthers previous work initiated toward the development of a common ontology in this area.**

Index Terms Collective Intelligence, Shared vocabulary, Ontology, Emergency Management

I. INTRODUCTION

Shared information flows improve and enable the efficiency of operations [1]. The Internet and web based communication technologies, constitute a powerful real time and distributed platform potentially well suited to speed the aggregation and exchange of large volumes of data, but despite the widespread availability of such platforms, large scale communication and coordination failures systematically characterize global scale emergency response operations [2]. These failures are caused by complex mixture of historical, political and cultural conflicts that are beyond the scope of IT, however misaligned information models are a well identified contributing factor [3].

II. COLLECTIVE INTELLIGENCE

The widespread adoption of web based platforms is facilitating the emergence of 'collective intelligence' , whereby autonomous distributed agents, contribute in near real time data, information, and 'intelligence' about a particular event, are considered a emerging field of potential great application for emergency management. In the definition adopted in this paper collective intelligence is: *the dynamic aggregation of cognitive, reasoning and knowledge resources of humans supported by intelligent and networked information systems* [6]. In order to accommodate for such polymorphism (many facets) Collective Intelligence can be viewed as a type of 'complex system', where "Complex systems do not have a centralizing authority and are not designed from a known specification, but instead involve disparate stakeholders creating systems that are functional for other purposes and are only brought together in the complex system because the individual "agents" of the system see such cooperation as being beneficial for them." Some of the key properties of complex systems are :

Emergence resulting from dynamic combination of a system components, and based on the dependence of the whole on its parts, their parts mutual interdependence and specialization.

Pattern formation: the visible, orderly outcomes of self-organisation and the common principles behind matching behaviours.

Paradoxes Diverse and heterogeneous components of a system can results in contrasting and sometimes opposed characteristics both being present , such as simplicity and complexity, order and disorder, random and predictable behaviour

In the context of emergency management, CI offers the possibility to harness fast paced, virtual interactions and information exchanges that take place thanks to the availability of web based IT tools. Not only collective intelligence allows us to harness knowledge and brainpower of other individuals, but also it makes use of advanced functional capabilities provided by contemporary software tools, that can enhance human reasoning and analytical capabilities, all in a connected real time environment such as the web. Examples of occurrences of collective intelligence during emergencies are documented in literature, both academic and from practitioners accounts [4]. A salient characteristic of collective intelligence is **self organization**, also known as *autopoiesis* or *autocatalysis,* which is the ability of individual entity or component within an ecosystem to develop internal and external dynamics that provide self synchronization within the other entities or parts of system. In the study 'Fundamental concepts of collective intelligence' based on the observation of social insects [5], researcher William Sulis derives general principles of collective intelligence, that can, at least in part be used to understand human collective intelligence. These are:

1) *Stochastic Determinism (SD):* Stochastic, from the Greek "stochos" means "aim, guess", referring to conjecture and randomness. SD refers to the global logic that underpins the development of a system as the result of individual behaviours of a community of individuals appear to be random, or at least not following a hierarchical, centrally imparted behaviour, yet resulting in an organic, socially purposeful action. In human CI systems, this principle corresponds to the state of randomness of communities where participants are not selected on the merit of their seniority, rank or expertise, but wider participation is encouraged. Collective intelligence relies on the principle of open participation, characterized by chaotic patterns of interaction.

2) *Interactive Determinism (ID),* according to this principle, the interaction among the constituents of a system results in some unique collective property, a type of synergy where the sum is more than just the sum of its factors. Thanks to ID, a system defines it dynamic processes on the fly, as a constant flow of chain reactions that are 'unpredictable' however they follow some built in logic. Self organization is the result of ongoing interactive determination and adjustments

3) *Non representational, Contextual Dependence.* According to this principle, collective intelligent behavior is determined by adaptive responses to the interaction among individuals and their environment, and does not depend on a shared cognitive representation (knowledge representation, model of the world). Biological systems do not have mental capabilities in the cognitive sense, and their behavior can be boiled down to a set of environmental responses. In biology, simple life forms do not posses the cognitive apparatus to support mental capabilities required for mental models to form, in contrast to human systems whose communication depends on shared conceptual and semantic models. Unlike lower order biological systems (entities with limited cognitive faculties) communication and information exchange in human systems require shared mental models, we'll discuss these in a separate paragraph.

In addition to the above principles, additional essential factors that enable and support collective intelligence of human systems, are diversity, autonomy, and accessibility of information, discussed in detail elsewhere [6].

III. SHARED MENTAL MODELS

The 'non representational contextual dependence principle' discussed above, shows how elementary biological entities where the principles of 'natural' collective intelligence are well established to enable cooperation, do not rely on explicit knowledge conceptualizations, but rather on sensorial inptus. Human communication however relies heavily on shared mental models for information understanding and utilization, and is necessary for operations coordination. Sharing common understanding of the structure of information among people or software agents is one of the more common goals in developing ontologies. An example provided by Noy and McGuinnes [7] says "suppose several different Web sites contain medical information or provide medical e-commerce services. If these Web sites share and publish the same underlying ontology of the terms they all use, then computer agents can extract and aggregate information from these different sites. The agents can use this aggregated information to answer user queries or as input data to other applications". To enable collective intelligence to take place, and to support 'distributed cooperation models', the underlying knowledge and information sharing

systems should be designed to conform to certain characteristics. Among the expected operational efficiencies that can be increased via distributed information systems [6] are:

- networked effects
- high transaction rates
- rapid adaptation to dynamic conditions
- flexibility
- the generation of "organic intelligence"
- the capability to harness 'global' intelligence

IV. DISTRIBUTED AND NETWORK ORIENTED INFORMATION MODELS

Even traditional and hierarchical command and control structures today are evolving to adapt to network oriented modes of operation [8], with implications in terms of social, technical and organisational changes which are impacting and revolutionizing the relation between IT and operations management in all sectors. This shift has started to become visible also in emergency response. At the heart of networked centric operations, lie open and distributed information models. Developing a common ontology that can be used and referenced to facilitate the communication and information interchange in support of large scale, distributed (non constrained to a single geography localised) emergency efforts, is an ambitious effort, first and foremost due to scale and magnitude of the project, and to its complexity. Initiatives aimed at the creation of a common information schemas in the emergency sector have already matured, and some are well under development, for example:

- Oasis Edxl [9]
- Cap [10]
- Australian All Hazard Taxonomy [11]

The W3C EIIF incubator, [12] has been tasked with doing preliminary evaluation and analysis and to issue suggestions for future work to be carried out by the W3C, (www consortium) in this direction. The work in this research paper is in part motivated by knowledge generated by the incubator, which could lead to a workgroup. W3C workgroups recommendations are considered 'prescriptive' by the web community, although they are not always free from controversy and constantly evolving, however they generally are accepted at a minimum as valid guidelines, and provide a set of formalised references that should be reasonably easy to adopt. The long term output of a W3C workgroup for emergency management, is expected to result in metamodel framework, in support of integrated information exchange during emergencies.

V. DYNAMIC MODELS, AND SCHEMA MATCHING

The central problems of 'application integration' and 'knowledge sharing and reuse' within the enterprise, are

caused primarily by lack of a consistent shared view that can be referenced uniformly and dynamically by different agents. According to John Sowa, sharing knowledge is problematic at three different levels: Knowledge, Ontology and Computation [13]. To bridge conceptual and semantic (knowledge) and representation gaps between different models, or between different aspects of a model, researchers have devised 'ontology matching' techniques such as mapping, aligning and merging of ontologies. A classification of matching techniques proposed by Shvaiko and Euzenat [14] proposes the distinction between approximate and exact techniques at schema level, and syntactic, semantic, and external techniques at element and structure level. The alignment of vocabularies and thesauri constitutes integral part of ontology matching and to any information model integration.

VI.   THE EIIF SHARED VOCABULARY

The lack of shared vocabulary is acknowledged as one of the causes 'semantic disconnectedness' on the web, and is a common, major problem in all sectors. Using different terms, definitions and concepts is one of the central causes of lack of semantic integration and divergence, therefore one of major obstacles to leveraging synergy and allowing collective intelligence to be catalyzed. Lexical and semantic distance may arise from differences among

1. Terms used by different agencies in the same operational field
2. Terms used by different agencies in different operational fields
3. Terms used by agencies in different countries, across 1. and 2. above

The output of the EIIF at this stage, consists of the creation of a conceptual framework that can be used toward the development of an interchange vocabulary format in support of operations.
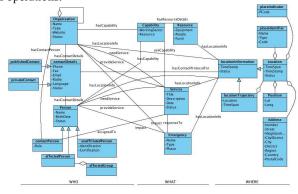

IMAGE: EIIF draft framework

This will include and rely on defining a set of 'context neu-

tral' terms that can be used as metadata and as the semantic backbone for exchanging dataset. A similitude can be drawn with the Dublin Core (http://dublincore.org). There is no single methodology to devise a working vocabulary for a given project or domain, although defining internal vocabularies – terms used by work group or project team – is generally a simpler task than defining a vocabulary for a wider community or sector, where conflicting interests, differing opinions and legacy views may contribute to the challenge. Vocabulary mapping techniques enable different vocabularies to be interoperable, and to be used in parallel, whereby a definition can be related to other definition simply by establishing and making explicit its lexical relation to other terms, see image 1.

Lexical relation	Example set	Underlying structure
synonym	A "happy" synonym set: {*happy, joyful, glad*}	simple set
measurement	A "temperature" set: {*cold, cool, lukewarm, warm, hot*}	scale
opposite	A "social relation" set: {(*student, teacher*), (*patient, doctor*)}	set of pairs
generic-specific whole-part	A "whole-part" tree: ○ *house* ○ ■ *roof* ■ *walls* ■ *floor*	tree

IMAGE 1 (source: http://www.sil.org/linguistics)

Vocabulary mapping and alignment is relatively simpler than a full conceptual and ontological mapping and alignment. Terms in vocabulary sets merely represent and point to concepts, that are further specified by ontologies in terms of axioms and rules. They do however represent a first, essential and necessary step in the process toward ontological and information model alignment. The fundamental lexical relations are hyperonymy, hyponymy, synonymy, antonymy, holonymy, meronimy [15] Identifying the lexical relations between terms and concepts can expose inherent semantic structures in the language in use. This particular vocabulary definition and mapping task being undertaken as part of the EIIF incubator aims to:

1. clarify the meaning of terms used by the incubator group itself in its documents
2. compare and relate these terms to terms used by agencies operating the field
3. refine and justify the lexical choices being made by the incubator
4. lay the foundations toward a possible unified interchange vocabulary, or emergency management set of metadata

The first portion of the task is relatively trivial, and consists of creating explicit definitions for the words in use, similar to creating glossary entries, or entries for a data dictionaries in software projects. Some are standard terms in use (fax, telephone, contact) and therefore have a generic definition which is already applicable, but some could also have a model specific definition within the given context of the framework, which needs to be specified (example: point of contact). The rationale for opting for novel terminology, as opposed to adopting terms already in use by established agencies in the field, has been questioned, and needs to be justified. The selection of terminology leads directly to the heart of the ontological problem being tackled. Terms in use come with a cultural and operational legacy that may be peculiar to a sector, an organisation, an industry, or a country, and which may constitute an inherent barrier to communication, as well as facing resistance to acceptance and adoption by others, especially the wider community. How the choice of vocabulary directly influences the appropriateness of an information model is studied elsewhere [17]. Modern operation model design favours the adoption of neutral terms that carry as little implications as possible. The work of the EIIF, in particular the efforts toward common vocabulary development, explored this direction.

Example

One of the documented disputed terms in emergency information systems is ' victim'. Victim is a widespread English language term in use in emergency management operations worldwide. There is cursory evidence that some people affected by adverse events resent being called 'victim', as this conveys an image of passivity, helplessness and impotence. While many would agree that people impacted by adversity and in need of emergency aid have higher priorities than disputing preferred naming conventions, it could be argued that 'victim' in itself does is not necessarily a meaningful word (dead or alive, injured or healthy, having lost someone or something etc). Terminological correctness and sensitivity is not only desirable where possible, but also potentially easier to map to vocabularies existing in use. The term 'victim' is therefore semantically mapped to a preferred context neutral term, such as 'person affected by event', corresponding to the current naming convention for this entity in the EIIF draft.

Semantic cluster identifier: person
Includes: (victim)
            recipient of
            beneficiary of aid
            client/patient/user
            Displaced Person

Similar processes apply for many terms used conventionally by the emergency management sector, for example, 'disaster'. During open community discussion, it emerged that the term 'disaster' is not necessarily representative of the range of adverse events that constitute an emergency, and it may have some undesired connotations. Therefore

Semantic cluster identifier: event
Includes: (disaster)
            Occurrence
            Incident (used by OASIS)
            Emergency

And so on, for resource/supply, as well as other terms used as identifier for each of the entity/classes of the EIIF diagram. At the moment, methods are being considered to map the glossaries to each other, which are being devised with reference to similar work already carried out in this area, specifically in the medical domain. [16] A preliminary method outline consists of

a) building a concept library
b) associating terms from different glossaries to shared concept creating semantic clusters
c) measure the semantic distance of the terms from each other to determine their 'relatedness'
d) create synonym map based on semantic distance, and a retrieval mechanism that would allow users of term to 'see also' related terms, and in RDF/OWL to simply adopt the same_as relation.

Standard concept extraction, lexical analysis and semantic matching methods are being evaluated for consideration to be adopted in this project.

VII. VALIDATION OF THE APPROACH

Glossaries are being collected via web searches and thanks to community collaboration, and although a consolidated and exhaustive list is not yet available, a simple web search for 'emergency management glossary' yields a very long list of results, which have to be ranked for relevance and usefulness to the EIIF scope. A subset of 15 glossaries has already been made available by researchers working on a EU funded project [17]. A preliminary analysis of these glossaries shows great divergence in conceptual and information models being represented by the corresponding lists of terms depending on the operational and institutional focus, confirming and validating the proposed approach that semantic mapping and clustering of terms in use would be a useful, and even essential step, in the development of a common ontology for emergency management. Additional vocabularies are being contributed from different members of the group.

VIII. IMPLEMENTATION

Many possible implementation options are being considered. The first intuitive choice is to load the vocabulary/vocabularies a 'semantic media wiki' type of environments, which supports intuitive WYSIWYG editors that allow users to cut and paste their glossaries using interfaces that do not require specialised skills other than ordinary web based word processing. The environment currently being used for a preliminary aggregation of terms is the Knoodl platform (knoodl.com) that supports the automated export of vocabulary both to OWL, RDF and XSD schemas, as well

as the recently added function of supporting SPARQL queries via endpoint. Once a vocabulary is represented/exported to the preferred choice of schema, or to a schema that enables transformation between schemas, such as XSD, users will be able to view, query and manipulate the vocabularies at will, as a vocabulary implementation should be platform independent and should not constrain the user to adopt any particular technology. It is envisaged that intuitive web based platforms can be developed further to facilitate additional and more advanced functionalities in relation to the semantic integration and querying of vocabularies.

IX. PROBLEMS

Compiling a glossary of terms used in a model, or document, should be easy, but the exercise under construction has turned out not to be trivial at all. First, several semantic and conceptual gaps exist within the terms used by the framework, and defining a vocabulary is helping us to identify them. In addition, mapping the vocabularies to one another is exposing the same gaps in the terminology used in the sector although this information can be used to discover knowledge areas. The lack of a common presentation latyer and publication format (some are published ad PDFs, some as word files, some are encoded in some software) is not helping their consultation and mapping, so we are compiling a tabular version, i.e, listing them all in a table. Where each term is compared across several vocabularies, where each vocabulary is a column. See example below.

	VOCAB1	VOCAB2	VOCAB 3	
TERM1				
TERM2				
TERM3				

However we are still exploring techniques to represent semantic distance more sophisticated and faster than using tables. Another problem is that different vocabularies are encoded with different schemas (syntactically), and last but not least, that the variety of relations between the various terms and their meanings is not entirely supported by the existing semantic frameworks and languages, such as RDF/OWL, or if they are, the documentation in support of such techniques is not sufficiently extensive nor detailed for a casual user. Further research is being undertaken to explore solutions to these problems.

X. CONCLUSION AND FUTURE WORK

This paper is a work in progress report. Shared vocabularies are at the heart of knowledge integration This is true for information and operational models, for semantic web, web 2.0 applications and environments, and beyond. Sharing vocabularies does not necessary mean that stakeholders must universally agree on preferred lexical choices, as lexical freedom and diversity are in important factors in cultural identity, for individuals as well as for organisations. Shared vocabularies however rely on rigorously explicit terms and definitions adopted in language, be it natural, programming or semi automated, and on representing these definitions using commonly agreed, declared and publicly accessible schemas to facilitate and support information exchange. In this paper we discuss the centrality of shared vocabulary approaches to enable collective intelligence in the emergency management sector, and illustrate a working example of how such lexical and conceptual heterogeneity can be reduced with the development of a shared vocabulary. We have also discussed the challenges and problems faced during the early stages of our research. The work ahead includes the systematic study and analysis of internal workgroup glossary structure, a comparative analysis of our elected terms with other existing glossaries in use and techniques to the improvement and refinement of mapping and matching techniques. The final goal of this project is to contribute to efforts toward web integration using natural languag, semantic technologies.

REFERENCES

[1] D'Amoursa, Montreuila, Lefrançoisa Soumis, "Networked manufacturing: The impact of information sharing" International Journal of Production Economics Volume 58, Issue 1, 5 January 1999, Pages 63-79

[2] Peter S Anderson "Information technology : where is it in the coordination of emergency services?" Police technology Asia Pacific police technology conference proceedings of a conference held 12-14 Nov 1991

[3] Chen Sharman Rao Upadhyaya "Emergency Response Coordination and IT support: Contingency and Strategies" State Univ. of N Y at Buffalo

[4] Vieweg, Palen, Liu, "Collective Intelligence in Disaster: Examination of the Phenomenon in the Aftermath of the 2007 Virginia Tech Shooting" ConnectiveIT & the Natural Hazards Center University of Col., Boulder

[5] W. Sulis "Fundamental Concepts of Collective Intelligence"

[6] P. Di Maio "Making sense of collective intelligence" Cutter Consulting 2008

[7] Noy McGuinness "Guide to Create your First Ontology"

[8] Lee, Bae Han, Kim, and Park "Network oriented Network-centric Feature-based Modeling" Concurrent Engineering Research Electronics and Telecommunications Research Institute (ETRI) South Korea

[9] EDXL xml.coverpages.org/edxl.html

[10] CAP http://www.oasis-open.org/

[11] T. Hunter "All Hazards Taxonomy"

[12] EIIF W3C http://www.w3.org/2005/Incubator/eiif/

[13] J. Sowa http://www.jfsowa.com/ontology/ontoshar.htm

[14] Euzenat Shvaiko "Survey of Schema based Matching approaches"

[15] M Krifka "Lexical Relations" Lexical Semantics, Fall 1998

[16] "A Term-mapping Framework for Data Mining in Heterogeneous Medical Data Sources"

[17] Prof A. David, Universita di Firenze Private correspondence,

[18] M. Stede "Lexical Choice Criteria in Language Generation"

[19]Department of Computer Science University of Toronto Canada